Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome sequence data of multidrug-resistant *Pseudomonas aeruginosa* WO7 from a hospital wastewater treatment plant in Thailand

Montri Yasawong [a,b], Thunwarat Songngamsuk [a], Manassanan Phatcharaharikarn [a], Pichapak Sriyapai [c], Kun Silprasit [d], Arin Ngamniyom [d], Thayat Sriyapai [d,*]

[a] *Program on Environmental Toxicology, Chulabhorn Graduate Institute, Bangkok 10210, Thailand*
[b] *Center of Excellence on Environmental Health and Toxicology (EHT), OPS, MHESI, Bangkok 10400, Thailand*
[c] *Department of Microbiology, Faculty of Science, Srinakharinwirot University, Bangkok 10110, Thailand*
[d] *Faculty of Environmental Culture and Ecotourism, Srinakharinwirot University, Bangkok 10110, Thailand*

ARTICLE INFO

Article history:
Received 24 February 2024
Revised 12 March 2024
Accepted 13 March 2024
Available online 18 March 2024

Dataset link: Draft genome sequence data of multidrug-resistant Pseudomonas aeruginosa WO7 from a hospital wastewater treatment plant in Thailand (Original data)

Keywords:
Multidrug resistance
Fluoroquinolone
Fosfomycin
Sulphonamide
NDM-1
OXA-10
OXA-50
VEB-1

ABSTRACT

Multidrug-resistant *Pseudomonas aeruginosa* WO7 was isolated from an untreated water sample from a hospital wastewater treatment plant in Thailand. This report presents the draft genome sequence data of *P. aeruginosa* WO7. Genomic DNA was obtained from a pure culture of *P. aeruginosa* WO7, and paired-end reads were generated using an Illumina MiSeq sequencer. The draft genome consisted of 111 contigs with a total size of 6,784,206 base pairs, an N50 of 209,424 base pairs, and a GC content of 65.85%. The dDDH value between WO7 and *Pseudomonas aeruginosa* DSM 50071^{T} was determined to be 90.7%, indicating that the strain is *Pseudomonas aeruginosa*. The data presented indicate the potential for bacterial classification, comparative genomics, comprehensive analysis of antimicrobial resistance, and assessment of bacterial virulence factors in *P. aeruginosa*. The draft genome sequence data have been deposited at the NCBI under Bioproject accession number PRJNA550309.

* Corresponding author.
E-mail address: thayat@g.swu.ac.th (T. Sriyapai).

https://doi.org/10.1016/j.dib.2024.110344
2352-3409/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/)

© 2024 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY-NC license
(http://creativecommons.org/licenses/by-nc/4.0/)

Specifications Table

Subject	Biological sciences
Specific subject area	Omics: Genomics
Data format	Raw and analysed
Type of data	Tables, figures
Data collection	DNA was extracted using the PureLink^{TM} Genomic DNA Mini Kit and sequenced using an Illumina MiSeq. Fastp v0.23.4 was used for adapter trimming and quality filtering. Genome assembly was performed using Unicycler v0.5.0 and assembly metrics were determined using QUAST v5.0.2. Genome quality was assessed using CheckM v1.1.2. The phylogenomic tree and dDDH values were analysed using the Type (Strain) Genome Server. The genomic map was generated using Proksee. Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline. Identification of antimicrobial resistance genes and prediction of resistance phenotypes were performed using ResFinder v4.4.2.
Data source location	*Pseudomonas aeruginosa* WO7 was isolated from sewage in a hospital wastewater treatment plant in Nakhon Nayok Province, Thailand (14°06′40″N, 100°59′03″E).
Data accessibility	Sequencing data were deposited in the National Center for Biotechnology Information (NCBI) Genbank database under the accession number JBAHYF000000000. The deposited draft genome sequencing data are available at https://www.ncbi.nlm.nih.gov/nuccore/JBAHYF000000000.

1. Value of the Data

- The draft genome data of *P. aeruginosa* WO7 has potential value for scientific studies in the fields of bacterial taxonomy and ecology, particularly for identifying and distributing taxa.
- The draft genome data of *P. aeruginosa* WO7 offers potential benefits for comparative genomic research on other *Pseudomonas* species with multidrug resistance.
- Elucidating the draft genome data of *P. aeruginosa* WO7 could assist in identifying antimicrobial resistance genes and predicting drug resistance phenotypes.

2. Background

Pseudomonas aeruginosa, a Gram-negative bacterium, is a major cause of nosocomial infections [1]. *P. aeruginosa* is commonly associated with respiratory, urinary, and wound infections and can cause bacteraemia [1,2]. *P. aeruginosa* is particularly problematic in hospitals, causing infections in patients on ventilators, with catheters, or with surgical wounds [1,2]. *P. aeruginosa* has developed resistance to many antibiotics, making infections difficult to treat [2,3]. In 2017, multidrug-resistant *P. aeruginosa* caused an estimated 32,600 infections in hospitalised patients and an estimated 2700 deaths in the United States [3]. The ability of *P. aeruginosa* to develop resistance to antibiotics is a major concern in healthcare [3].

3. Data Description

Here we present the draft genome sequence data of *P. aeruginosa* WO7, including its gene clusters with antimicrobial resistance genes and predicted phenotypes (Fig. 1).

The genome is composed of 111 contigs with a total size of 6,784,206 bp, an N50 value of 209,424 bp, and a GC content of 65.85% (Table 1).

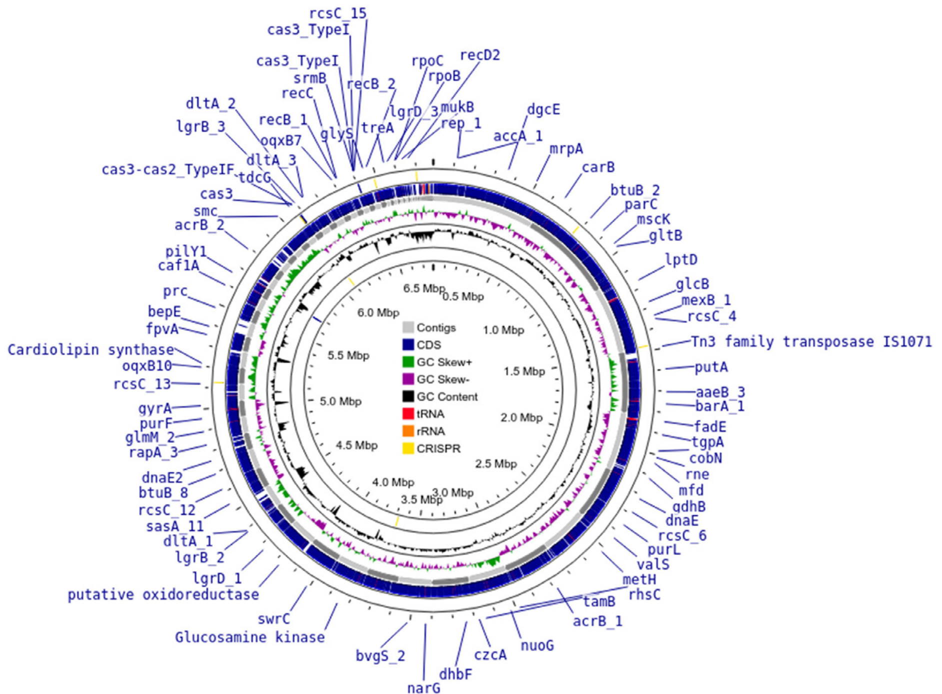

Fig. 1. The genome map of *P. aeruginosa* WO7 was generated using Proksee [4]. The coding DNA sequences (CDSs) are indicated by blue arrows, and the contigs are represented by grey arrows. GC skew+ and GC skew- are represented by green and purple peaks, respectively, and GC content is indicated by black peaks.

Table 1
Genomic features and assembly statistics for *P. aeruginosa* WO7.

Attribute	*P. aeruginosa* WO7
Genome size (bp)	6,784,206
Number of contigs	111
Genome coverage	44×
GC content (%)	65.85
Largest contig (bp)	577,113
N50	209,424
N75	91,658
L50	10
L75	23
Total gene	6,385
Total CDS	6,323
tRNA	54
rRNA	4
ncRNA	4
CRISPR repeat	3

The draft genome of *P. aeruginosa* WO7 is reported to be 99.68% complete, with an estimated contamination of less than 1%. The digital DNA-DNA hybridisation (dDDH) value between strain WO7 and *P. aeruginosa* DSM 50071^{T} was 90.7%. Fig. 2 shows the phylogenomic tree of strain WO7 and closely related type strains, confirming that strain WO7 is a *Pseudomonas aeruginosa* strain.

Whole genome sequence (WGS)-based antimicrobial susceptibility testing (AST) revealed the presence of known multidrug resistance genes with significant similarity, namely *aac(6′)-lb3*

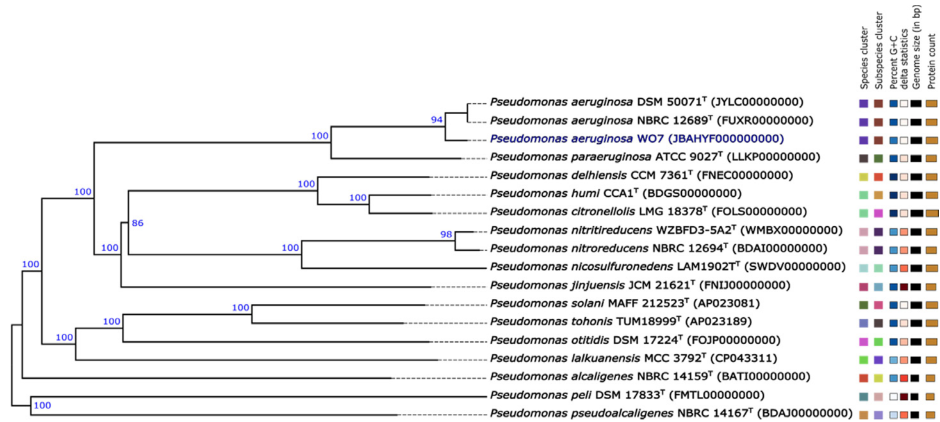

Fig. 2. The phylogenomic tree was reconstructed using whole-genome sequence data from *P. aeruginosa* WO7 and its closely related type strain on the TYGS platform. Branch numbers were determined on the basis of pseudo-bootstrap support values greater than 80% from 100 replicates using Genome Blast Distance Phylogeny (GBDP), with an average branch support of 95.7%.

(99.82%), *aadA1* (99.86%), *ant(4′)-llb* (99.47%), *aph(3)-IIb* (99.63%), bla_{NDM-1} (100.00%), bla_{OXA-10} (100.00%), bla_{OXA-50} (99.37%), bla_{PAO} (99.50%), bla_{VEB-1} (99.67%), *catB7* (98.58%), *cmlA1* (98.57%), *crpP* (98.48%), *fosA* (99.75%), *sul1* (99.82%), *tet*(A) (99.57%), and *tet*(G) (100.00%) in the *P. aeruginosa* WO7 genome (Table 2). Prediction of resistance phenotypes suggests that *P. aeruginosa* WO7 may be resistant to antibiotics from seven classes, including aminoglycoside, beta-lactam, phenicol, fluoroquinolone, fosfomycin, sulphonamide, and tetracycline (Table 2). The draft genome sequence data could facilitate a comprehensive analysis of the antimicrobial resistance and bacterial virulence factors of *P. aeruginosa*.

4. Experimental Design, Materials, and Methods

4.1. Bacterial Isolation

Untreated wastewater was collected from a treatment plant at a hospital (14°06′40"N, 100°59′03"E) in Nakhon Nayok Province, Thailand. Ten millilitres of untreated wastewater was subjected to 10-fold serial dilution using sodium chloride solution (0.85% (w/v) NaCl). Then, 100 µL of the mixture was spread onto Hicrome™ KPC agar (Himedia, India) and incubated at 37°C for 20 h. After selecting a single colony of strain WO7, it was subcultured for purification on tryptic soy agar (Himedia, India). Strain WO7 was cultured in tryptic soy broth (TSB) at 37°C with shaking at 250 rpm for 24 h. It was then stored for long-term preservation in TSB containing 25% glycerol at -80°C.

4.2. Genomic DNA Preparation

Genomic DNA (gDNA) was extracted from overnight cultures of strain WO7 grown in TSB using the PureLink™ Genomic DNA Mini Kit (Invitrogen, USA) according to the manufacturer's instructions. The quality of gDNA was assessed using agarose gel electrophoresis and NanoDrop spectrophotometry (Thermo Scientific, USA).

Table 2
Identification of antimicrobial resistance genes and prediction of resistance phenotypes from the WO7 genome sequence.

Resistance ene	Identity (%)	Alignment length/gene length	Position in the reference	Contig	Position in the contig	The resistance phenotype	Antibiotic class	GenBank accession number
aac(6')-Ib3	99.82	555/555	1-555	75	27-581	Amikacin, Tobramycin	Aminoglycoside	X60321
aadA1	99.86	722/792	1-722	94	207-928	Spectinomycin, Streptomycin		JQ414041
ant(4')-IIb	99.47	756/756	1-756	52	7,313-8,068	Amikacin, Tobramycin, and Isepamicin		AY114142
aph(3')-IIb	99.63	807/807	1-807	18	99,053-99,859	Kanamycin, Neomycin, Paromomycin, Ribostamycin, Butiromycin, Gentamicin, and Unknown aminoglycoside		CP006832
bla_{NDM-1}	100.00	813/813	1-813	52	14,679-15,491	Amoxicillin, Amoxicillin+Clavulanic acid, Ampicillin, Ampicillin+Clavulanic acid, Cefepime, Cefixime, Cefotaxime, Cefoxitin, Ceftazidime, Ertapenem, Imipenem, Meropenem, Piperacillin, Piperacillin+Tazobactam, Temocillin	Beta-lactam	FN396876
bla_{OXA-10}	100.00	610/801	1-610	96	1-610	Amoxicillin, Ampicillin, Aztreonam, Piperacillin, Piperacillin+Tazobactam		J03427
bla_{OXA-50}	99.37	789/789	1-789	12	112,364-113,152	Amoxicillin, Ampicillin		AY306130
bla_{PAO}	99.50	1,194/1,194	1-1,194	18	85,517-86,710	Amoxicillin, Ampicillin, Cefepime, Ceftazidime		AY083592
bla_{VEB-1}	99.67	900/900	1-900	64	3,213-4,109	Amoxicillin, Amoxicillin+Clavulanic acid, Ampicillin, Ampicillin+Clavulanic acid, Aztreonam, Cefotaxime, Cefoxitin, Cefepime, Ceftazidime, Piperacillin, Piperacillin+Tazobactam, Ticarcillin, Ticarcillin+Clavulanic acid		HM370393
catB7	98.59	639/639	1-639	20	51,720-52,358	Chloramphenicol	Phenicol	AF036933
cmlA1	98.57	1,260/1,260	1-1,260	75	796-2,055	Chloramphenicol		M64556
crpP	98.48	198/198	1-198	33	52,671-52,868	Ciprofloxacin	Fluoroquinolone	HM560971
fosA	99.75	408/408	1-408	1	414,986-415,393	Fosfomycin	Fosfomycin	ACWU01000146
sul1	99.82	556/867	1-556	93	1-556	Sulfamethoxazole	Sulphonamide	EU780013
tet(A)	99.57	1,174/1,200	1-1,173	64	1,289-2,461	Doxycycline, Tetracycline	Tetracycline	AY196695
tet(G)	100.00	1,176/1,176	1-1,176	52	2,811-3,986	Doxycycline, Tetracycline		AF133140

4.3. Whole Genome Sequencing and Assembly

Library preparation for DNA sequencing was performed using sparQ Frag & DNA Library Prep (QuantaBio, USA) with 100 ng of gDNA. gDNA was fragmented through an enzymatic reaction and subsequently purified using magnetic beads. Following this, an adaptor index was ligated to the fragmented DNA. The quantity and quality of the indexed libraries were measured using the 2100 Bioanalyzer (Agilent, USA) and QFX Fluorometer (DeNovix, USA), respectively. They were then pooled in equimolar amounts. Cluster generation and paired-end 2×250 nucleotide read sequencing were performed using an Illumina MiSeq sequencer. Quality assessments, adapter trimming, and quality filtering were performed using Fastp v0.23.4 with default settings [5]. *De novo* genome assembly was performed using the raw reads and Unicycler v0.5.0 with default settings [6]. The assessment of genome assembly metrics was performed using QUAST v5.0.2, employing the default parameters [7].

4.4. Taxonomic Identification of the Strain

The quality of the genome sequence was assessed using CheckM v1.1.2 with default settings [8]. An analysis of digital DNA-DNA hybridisation (dDDH) and a phylogenomic tree based on the whole genome sequences of WO7 and its related type strains was conducted using the Type (Strain) Genome Server (TYGS) [9].

4.5. Genome Annotation and Antimicrobial Gene Prediction

Proksee [4] generated a genomic map of WO7, and the genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) with default settings [10]. Furthermore, ResFinder v4.4.2 with default settings [11] was used to perform whole genome sequence (WGS)-based antimicrobial susceptibility testing.

Limitations

Next-generation sequencing techniques generate vast amounts of data. However, the resulting *de novo* genome assemblies often lack completeness. These assemblies may contain shortcomings that make them vulnerable to annotation errors, particularly in the imprecise estimation of genes that may exist in the draft genome of WO7.

Ethics Statement

This study did not involve human or animal subjects. The authors declare that this manuscript is original and has not been published elsewhere.

Data Availability

Draft genome sequence data of multidrug-resistant Pseudomonas aeruginosa WO7 from a hospital wastewater treatment plant in Thailand (Original data) (Genbank).

CRediT Author Statement

Montri Yasawong: Methodology, Data curation, Writing – original draft, Writing – review & editing; **Thunwarat Songngamsuk:** Methodology, Data curation; **Manassanan Phatchara-**

harikarn: Methodology, Data curation; **Pichapak Sriyapai:** Methodology, Data curation; **Kun Silprasit:** Methodology, Data curation; **Arin Ngamniyom:** Methodology, Data curation; **Thayat Sriyapai:** Methodology, Data curation, Writing – original draft, Writing – review & editing, Supervision.

Acknowledgments

This research was supported by the 2022 research fund of Srinakharinwirot University grant no. 014/2565, the Thailand Science Research and Innovation project code 180874 (FRB660044/0240), and the Center of Excellence on Environmental Health and Toxicology (EHT), OPS, Ministry of Higher Education, Science, Research and Innovation.

Declaration of Competing Interest

The authors declare that they have no conflicts of interest that could have influenced the work reported in this paper.

References

[1] D. Reynolds, M. Kollef, The epidemiology and pathogenesis and treatment of *Pseudomonas aeruginosa* infections: An update, Drugs 81 (2021) 2117–2131, doi:10.1007/s40265-021-01635-6.
[2] M. Larrosa, P. Truchado, J.C. Espín, F.A. Tomás-Barberán, A. Allende, M.T. García-Conesa, Evaluation of *Pseudomonas aeruginosa* (PAO1) adhesion to human alveolar epithelial cells A549 using SYTO 9 dye, Mol Cell Probes 26 (2012) 121–126, doi:10.1016/j.mcp.2012.03.001.
[3] CDC's 2019 Antibiotic Resistance Threats Report, Centers for disease control and prevention, 2019 https://www.cdc.gov/drugresistance/pdf/threats-report/pseudomonas-aeruginosa-508.pdf. Accessed February 24, 2024.
[4] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C.Y. Chen, M. Graham, G. Van Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, Nucl. Acids Res. 51 (2023) W484–W492, doi:10.1093/nar/gkad326.
[5] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ pre-processor, Bioinformatics 34 (2018) i884–i890, doi:10.1093/bioinformatics/bty560.
[6] R.R. Wick, L.M. Judd, C.L. Gorrie, K.E. Holt, Unicycler: resolving bacterial genome assemblies from short and long sequencing reads, PLoS Comput. Biol. 13 (2017) e1005595, doi:10.1371/journal.pcbi.1005595.
[7] A. Mikheenko, A. Prjibelski, V. Saveliev, D. Antipov, A. Gurevich, Versatile genome assembly evaluation with QUAST-LG, Bioinformatics 34 (2018) i142–i150, doi:10.1093/bioinformatics/bty266.
[8] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. 25 (2015) 1043–1055, doi:10.1101/gr.186072.114.
[9] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, Nat. Commun. 10 (2019) 2182, doi:10.1038/s41467-019-10210-3.
[10] T. Tatusova, M. DiCuccio, A. Badretdin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucleic Acids Res. 44 (2016) 6614–6624, doi:10.1093/nar/gkw569.
[11] V. Bortolaia, R.S. Kaas, E. Ruppe, M.C. Roberts, S. Schwarz, V. Cattoir, A. Philippon, R.L. Allesoe, A.R. Rebelo, A.F. Florensa, L. Fagelhauer, T. Chakraborty, B. Neumann, G. Werner, J.K. Bender, K. Stingl, M. Nguyen, J. Coppens, B.B. Xavier, S. Malhotra-Kumar, H. Westh, M. Pinholt, M.F. Anjum, N.A. Duggett, I. Kempf, S. Nykäsenoja, S. Olkkola, K. Wieczorek, A. Amaro, L. Clemente, J. Mossong, S. Losch, C. Ragimbeau, O. Lund, F.M. Aarestrup, ResFinder 4.0 for predictions of phenotypes from genotypes, J. Antimicrob. Chemother. 12 (2020) 3491–3500, doi:10.1093/jac/dkaa345.